

EEXCESS Redesign

Mathias Möller,
`moellerm@fim.uni-passau.de`

Universität Passau, 94032 Passau

1 Motivation

Ziel dieser Arbeit ist es, ein Chrome Plugin zu entwickeln, welches, ähnlich zum EEXCESS Plugin, dem Benutzer beim Lesen einer Webseite weiterführende Quellen zum Thema der Seite vorschlägt. Dies soll es dem Nutzer ermöglichen, ohne weiteren Suchaufwand, weiterführende wissenschaftliche Quellen zu finden und so die Recherche zu vereinfachen und zu optimieren.

Während EEXCESS dem Nutzer nur Quellen zum Thema der gesamten Seite vorschlägt und diese am Rand der Browsers anzeigt, soll das neue Plugin die Paragraphen der Seite erkennen und analysieren und auf Basis dieser Informationen eine Anfrage an die Europeana-Datenbank schicken. Die Ergebnisse sollen dem Benutzer dann erkennbar zum Paragraphen gehörend angezeigt werden. Daraufhin soll der Benutzer die Möglichkeit haben, die Suchanfrage anzupassen (zum Beispiel durch Löschen/Hinzufügen von Suchwörtern).

Durch diese Änderungen im Design und Verhalten des Plugins sollen die Probleme des EEXCESS Plugins, die bei einer Nutzer-Evaluierung zu Tage gefördert wurden, behoben werden und die Usability des Programms verbessert werden.

Um dieses Ziel zu erreichen, wird die Arbeit in folgende Schritte unterteilt: Anzeige der Ergebnisse, Erklärung der Such-Anfragen Generierung, Anpassung der Suchanfrage durch den Benutzer und eventuell Ausarbeitung eines Konzepts zur automatischen Anpassung der Suchanfragen durch maschinelles Lernen.

2 Vorarbeit

Zu Beginn der Arbeit wird untersucht, wie die Suchanfrage an die Europeana-Datenbank aus den Paragraphen extrahiert werden kann und wie die Ergebnisse dem Benutzer dargestellt werden müssen um seine Arbeitsweise bestmöglich zu unterstützen.

3 Benutzeroberfläche

Um den groben Aufbau des Plugins zu skizzieren (ohne sich auf das Design festzulegen) wurden Mockups der Benutzeroberfläche angefertigt.

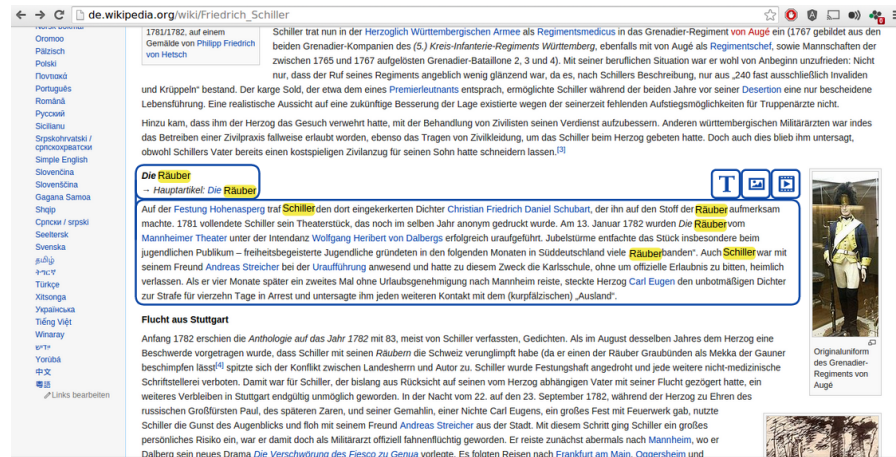


Abb. 1. Der erkannte Paragraph wird hervorgehoben und Buttons zum Auswahl der gewünschten Quellen-Art werden angezeigt

Während des Lesens einer Webseite wird das Plugin versuchen die Paragraphen-Struktur der Seite zu erkennen. Gefundene Paragraphen werden dann hervorgehoben (siehe Abb. 1) und eine Analyse des Inhalts wird durchgeführt. Mit den Ergebnissen dieser Analyse wird daraufhin eine Anfrage an die Europeana-Datenbank abgeschickt. Die benutzten Suchwörter werden im Text hervorgehoben, um es dem Benutzer zu Ermöglichen, neue Suchwörter per Klick hinzuzufügen oder andere zu Löschen.

Der Benutzer hat dann die Möglichkeit die Art der Quellen zu Bestimmen die ihm angezeigt werden sollen (siehe Abb. 2). Denkbare Auswahlmöglichkeiten wären Text-, Bild- oder Videoquellen.

4 Verwendete Technologien

Technologien zur Paragraphen-Erkennung sowie zur Textanalyse der Paragraphen werden vom Lehrstuhl für Medieninformatik bereitgestellt. Die Paragraphen-Erkennung wird gegebenenfalls optimiert um auch auf weniger strukturierten Seiten eine möglichst gute Erkennung zu ermöglichen.

Das Plugin wird seine Suchanfragen an die REST-API der Europeana-Datenbank schicken.

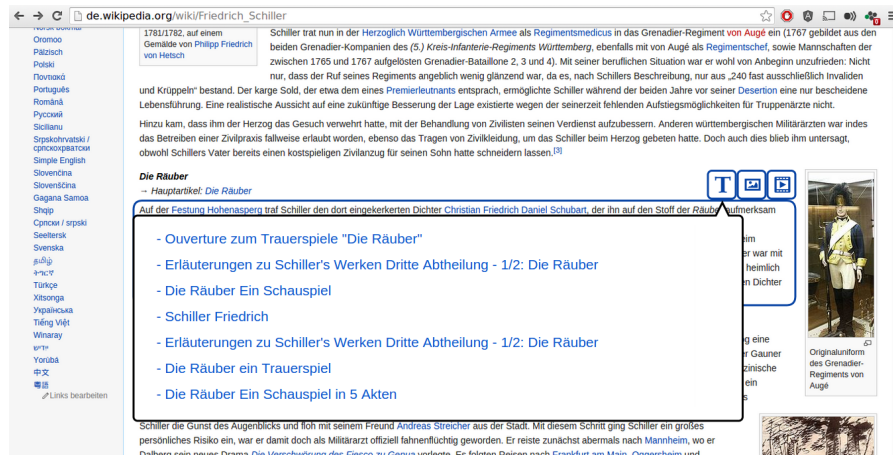


Abb. 2. Nach Auswahl der gewünschten Quellen-Art werden dem Benutzer die gefundenen Quellen angezeigt

5 Zeitlicher Ablauf der Ausarbeitung

- I) Recherche zu technischen Details der Umsetzung und Implementierung
- II) Skizzieren der Programmstruktur, REST-Anfragen, Interfaces
- III) Praktischer Teil:
 - Milestone 1:** Integration von Paragraphen-Erkennung in simples User-Interface
 - Milestone 2:** Integration der Textanalyse und senden der Such-Anfragen an Europeana-Datenbank
 - Milestone 3:** Anzeigen der Suchergebnisse
 - Milestone 4:** Manipulation der Suchanfrage
 - Milestone 5:** Design Ausarbeitung
 - Milestone 6:** (optional) Usability Tests
 - Milestone 7:** Endgültiges Design
- IV) Theoretischer Teil:
 - Einleitung/Motivation** Beschreibung der Motivation der Arbeit (Verbesserung des EEXCESS Plugins), Unterschiede zum EEXCESS Plugin
 - Related Work** Vergleich des Plugins und der verwendeten Technologien (Paragraph-Erkennung, Text-Analyse) mit ähnlichen Lösungen
 - Konzept**
 - Allgemeines und Verweis auf 4 Unterziele
 - Anzeige der Ergebnisse
 - Erklärung der Such-Anfragen Generierung
 - Anpassung der Suchanfrage durch den Nutzer
 - Verbesserung der Suchanfrage durch maschinelles Lernen
 - Implementierung** Beschreibung der REST-Anfragen, der verwendeten Technologien und der Vorgehensweise während der Implementierung

Evaluierung (optional) Usability Tests mit anschließendem Vergleich
mit den Ergebnissen der EEXCESS Evaluierung
Future Work Implementierung der automatischen Suchanfragen-Verbesserung
durch maschinelles Lernen